Name : .....

Roll No. : .....

Invigilator's Signature : .....

CS/M.Tech(MBIN)/SEM-1/MBIN-103/2012-13

2012
COMPUTATIONAL BIOLOGY -I

Time Allotted : 3 Hours *Full Marks :* 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Answer any *ten* questions from **Group A** and

four questions from **Group B**

GROUP – A

(Multiple Choice Type Questions)

1. Choose the correct alternatives for any *ten* of the following : 10 × 1 = 10

   i) Which of the following statements about secondary databases (like Prosite, Prints, Block) is correct ?

      a) They only contain secondary structural elements experimentally found in proteins

      b) They contain secondary structural elements experimentally found in proteins with their associated primary sequence

      c) The contain motifs (or homologous sequences) defined through multiple sequences alignment

      d) They contain motifs (or homologous sequences) defined through multiple alignments of secondary structural elements.

ii) The CATH classification is based on Class, Architecture, Topology, Homology. Which one of the following definitions is NOT correct ?

a) Class : 4 different domain classes : mainly α, mainly β, $\alpha - \beta$ (α/β and $\alpha + \beta$), low secondary structure content

b) Architecture : describes arrangement of secondary structure ignoring connectivities

c) Topoloty : describes the connectivity of secondary structure

d) Homology : groups domains that share > 70% of identities.

iii) The most widely used kind of scoring matrices for protein sequence alignments (e.g. BLOSUM matrices) was derived based on

a) Genetic code

b) Physico-chemical properties

c) Simple identity

d) Statistics of observed substitutions in "safe" multiple sequence alignments.

iv) Codon bias is a tool used in genomics to

a) identify open reading frames

b) differentiate between eukaryotic and prokaryotic DNA sequences

c) find regulatory sequences

d) identify a gene's function.

v) Which one of the following can be used for drawing a phylogenetic tree ?

a) MAUVE b) MEGA

c) Codon W d) Auto Dock.

vi) Maximum Parsimony model for phylogenetic analysis is a

a) Character based method

b) Distance based method

c) Pattern based method

d) Secondary structure based method.

vii) If you want literature information what is the best website to visit ?

a) OMIM b) Entrez

c) PubMed d) PROSITE.

viii) Which of the following amino acids is least mutable according to PAM matrix ?

a) Alanine b) Glutamine

c) Methionine d) Cysteine.

ix) You have two distantly related proteins. Which BLOSUM or PAM matrix is best to compare them ?

a) BLOSUM45 or PAM250

b) BLOSUM45 or PAM1

c) BLOSUM80 or PAM250

d) BLOSUM80 or PAM1.

x) Which of the following statements is correct ?

a) Nucleotide query against nucleotide sequence database is done by blastp

b) Protein query against a translated nucleotide sequence database is done by blastp

c) Translated nucleotide query against a protein database is done by blastx

d) Protein query against a protein database is done by tblastn.

xi) Which of the following statements about multiple alignments is correct ?

a) It is not possible to define a dynamic programming algorithm to align more than two sequences

b) The guide tree in CLUSTAL W is produced using a distance matrix method

c) The guide tree in CLUSTAL W does not influence the final alignment

d) All three of the above statements.

xii) A PSI-BLAST search is most useful when you want to

a) Find the rat ortholog of a human protein

b) Extend a database search to find additional proteins

c) Extend a database search to find additional DNA sequence

d) Use a pattern or signature to extend a protein search.

GROUP – B

(Long Answer Type Questions)

Answer any *four* of the following. 4 × 15 = 60

2. a) Explain the term 'Protein secondary structure".

b) State one method for protein secondary structure prediction. How helices and strands can be predicted using that method ?

c) Write a short note of SCOP. 2 + (6 + 2) + 5

3. a) Differentiate between Phenetic approach and Cladistic approach.

   b) Explain the concept of 'Bootstrapping' in phylogenetic tree.

   c) Write with a suitable example the basic steps followed in Neighbour Joining method to draw a Phylogenetic Tree. 2 + 3 + 10

4. a) Discuss the importance and applications of comparative genome analysis.

   b) What are the utilities of MUMmer for large scale genome comparison ? 10 + 5

5. (i) What is BLAST ?

   (ii) Write down the basis steps of BLAST algorithm.

   (iii) Let us suppose in a BLAST search you got an E-value of about 2×10^{-10}. What does this E-value mean ?

   (iv) Discuss the parameters on which E-value depends.

   2 + 7 + 3 + 3

6. a) Define similarity, identity and homology. "Two sequences are 10% homologous". Explain the meaning of the statement.

   b) Define Orthologous and Paralogous sequences.

c) Using Needleman and Wunch algorithm align the following 2 sequences :

GAAT and GCA

Consider Match = 2

Mismatch = – 1

And gap penalty = – 2. 3 + 2 + 3 + 7

7. a) What is scoring matrix ? Write down its importance.

b) Write any two differences between PAM and BLOSUM.

c) What is the mathematical relationship between PAM1 and PAM40 ?

d) "One PAM is a unit of evolutionary divergence in which 1% of the amino acids have been changed. Therefore after 100 PAMS every amino acid will be different". Is the statement correct ? If not, then explain why ?

e) What is gap penalty ? Assume that you have made an alignment with gap creation penalty 10 and gap extension penalty 0·6. Next you realign the same sequences with gap creation penalty 250 and gap extension penalty 15. Did the alignment change ? If yes, explain why ? 4 + 2 + 2 + 3 + 4